Nucleic Acids Research, 2006, Vol. 34, Database issue ***D221–D226***
doi:10.1093/nar/gkj029

Nuclear Receptor Signaling Atlas (www.nursa.org): hyperlinking the nuclear receptor signaling community

Rainer B. Lanz*, Zeljko Jericevic, William J. Zuercher[2], Chris Watkins, David L. Steffen[1], Ronald Margolis[3] and Neil J. McKenna

Department of Molecular and Cellular Biology, M602 and [1]Department Molecular and Human Genetics, M620, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA, [2]Discovery Research Chemistry, GlaxoSmithKline, Five Moore Drive, NTH-M2127, RTP, NC 27709-3398, USA and [3]Division, of Diabetes, Endocrinology and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Democracy 2, Room 693, 6707 Democracy Blvd, Bethesda, MD 20892-5460, USA

Received August 15, 2005; Revised and Accepted September 22, 2005

ABSTRACT

The nuclear receptor signaling (NRS) field has generated a substantial body of information on nuclear receptors, their ligands and coregulators, with the ultimate goal of constructing coherent models of the biological and clinical significance of these molecules. As a component of the Nuclear Receptor Signaling Atlas (NURSA)—the development of a functional atlas of nuclear receptor biology—the NURSA Bioinformatics Resource is developing a strategy to organize and integrate legacy and future information on these molecules in a single web-based resource (www.nursa.org). This entails parallel efforts of (i) developing an appropriate software framework for handling datasets from NURSA laboratories and (ii) designing strategies for the curation and presentation of public data relevant to NRS. To illustrate our approach, we have described here in detail the development of a web-based interface for the NURSA quantitative PCR nuclear receptor expression dataset, incorporating bioinformatics analysis which provides novel perspectives on functional relationships between these molecules. We anticipate that the free and open access of the community to a platform for data mining and hypothesis generation strategies will be a significant contribution to the progress of research in this field.

INTRODUCTION

The field of nuclear receptors has made rapid progress over the last several decades, from the initial characterization of ligands and receptors, to the cloning of receptor cDNAs (1,2), to the discovery of an ever-increasing number of coregulatory proteins recruited by receptors to regulate target gene transcription (3,4). More recently, targeted gene deletions combined with transcriptomic analyses have taken steps towards building a picture of the complex developmental, physiological and metabolic networks in which nuclear receptors, their ligands and coregulators participate. A growing appreciation of the role of these molecules in tissue homeostasis, aging and a variety of disease states including diabetes, obesity and cancer, has made them the focal point of intense clinical and pharmaceutical interest.

While the development of large public genomic and proteomic database collections such as National Center of Biotechnology Information (NCBI), UniProt, EBI and others have facilitated this process, a significant deficit remains in the manner in which data generated by the field are integrated, analyzed and disseminated for the benefit of the entire community. Traditional hypothesis-driven research modes make it difficult—if not impossible—to compare between different datasets.

The Nuclear Receptor Signaling Atlas (NURSA) was designed by the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute on Aging, and the National Cancer Institute of the NIH to apply discovery-driven approaches to advance system-wide knowledge in nuclear

*To whom correspondence should be addressed for Data coordination, database schema and data analysis models. Tel: +1 713 798 6478; Fax: +1 713 790 1275; Email: rlanz@bcm.tmc.edu
Correspondence may also be addressed to Neil McKenna for Electronic publishing and community and journal interactions. Tel: +1 713 798 8568; Fax: +1 713 790 1275; Email: nmckenna@bcm.tmc.edu

receptor biology [for a primer on NURSA see Margolis *et al*. (5)]. The use of novel technologies and discovery-based high-throughput applications in NURSA are designed to catalyze traditional hypothesis-driven research by the wider community.

The NURSA Bioinformatics Resource plays a central role in the realization of the consortium's objectives by organizing these datasets in formats that permit presentation in useful and interactive ways. The objectives of our resource are 2-fold:

(i) to develop a comprehensive, freely-accessible community resource, dedicated to consolidating legacy and future data in the nuclear receptor field.
(ii) adopting consistent, sustainable standards in data annotation to facilitate sharing and comparative analyses of data in the field. Accordingly, we are developing a bioinformatics solution for data submission, storage, curation, presentation and mining, with the aim to encourage its adoption as a framework for future experimental strategies across the community.

NURSA SYSTEM ARCHITECTURE

In technical terms, the NURSA system architecture represents an enterprise class, highly scalable and robust '3-tier' architecture based on the J2EE standard. Raw data is stored according to a relational database schema we developed for the Oracle database management system that runs on Sun hardware hosted at Baylor College of Medicine's Bioinformatics Research Center. The application layer uses Oracle Application Server and other J2EE server environment components to provide business logic as well as interfaces to both the database and the presentation/client layer built in ColdFusion MX. In practical terms for the user, this architecture translates to a reliable foundation for data storage (the database), a flexible infrastructure for rapidly handling user queries (the business layer) and an intuitive, interactive interface (the presentation layer).

Key elements of the NURSA information solution that will be discussed in greater detail here are some components of the database system and our approach to website design that emphasizes the ability of the user to readily integrate distinct datasets and other information sources.

Data sources

The field of nuclear receptors is characterized by a broad compass, a diverse terminology and many intricate functional relationships, representing a substantial challenge for the resource in applying a solution for data integration and consolidation. The first step was to describe data and information with a structured vocabulary. This is represented by a network of relational tables designed using guidelines based on the normalization theory and the concept of constraints to enforce data integrity (6,7). In this database schema, nuclear receptors, ligands and coregulators form the three main molecule classes around which our database schema design is evolving. A strongly modular character of the developing system ensures that it retains sufficient flexibility to accommodate a variety of present and future projects.

Figure 1 illustrates the flow of information within NURSA. Data are deposited from the following resources:

- high quality experimental data from NURSA laboratories;
- external generic databases with broad genome coverage;

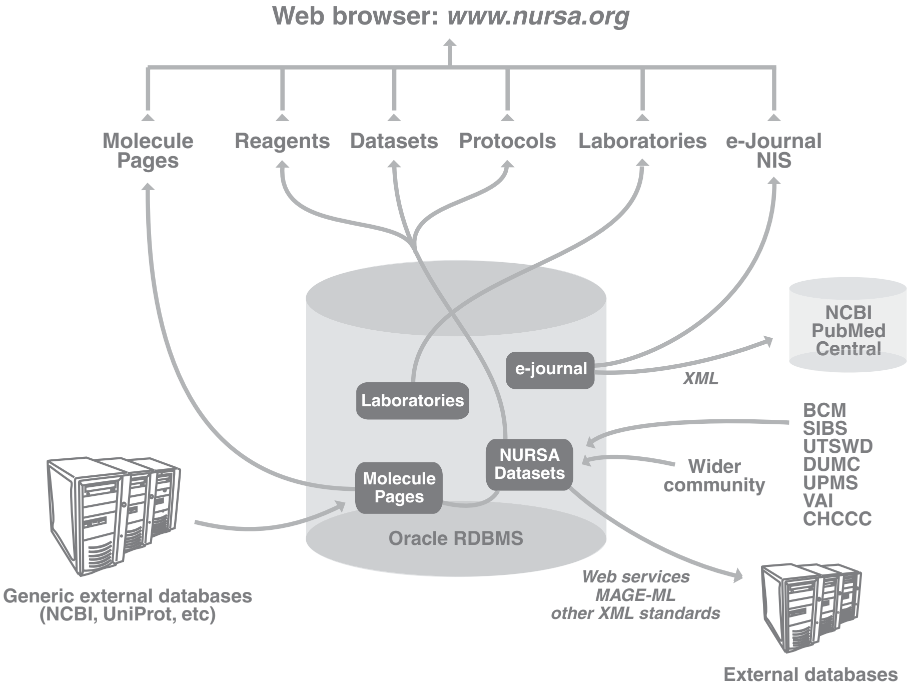
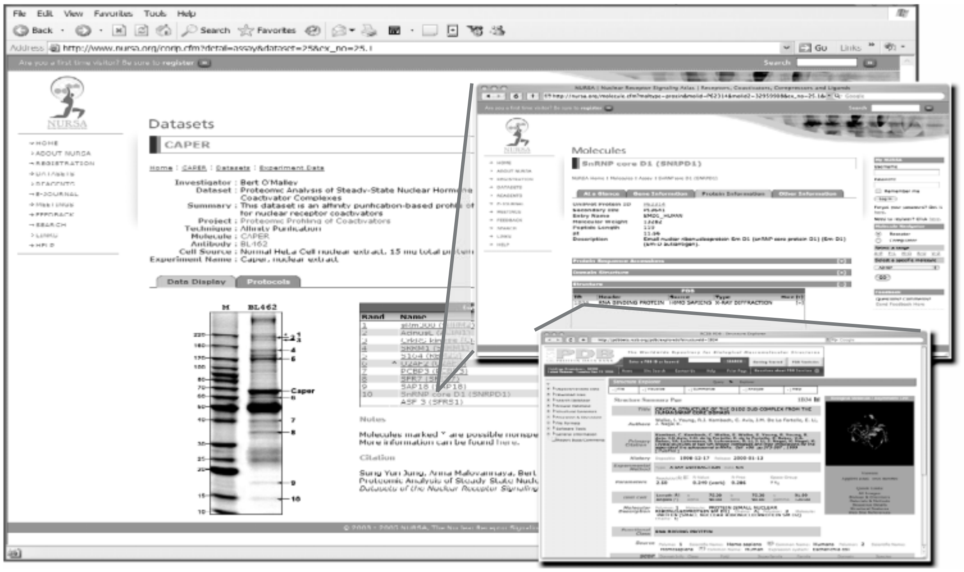


Figure 1. Data sources and information exchange of the NURSA information solution. Emphasis is given to the NURSA databases, the NURSA website organization and the information exchange between the NURSA databases and website, and with external databases. BCM, Baylor College of Medicine; SIBS, Salk Institute for Biological Studies; UTSWD, University of Texas Southwestern at Dallas; UPMS, University of Pennsylvania Medical School; DUMC, Duke University Medical Center; CHCCC, City of Hope Comprehensive Cancer Center; VAI, Van Andel Institute; RDBMS, relational database management system.



File Edit View Favorites Tools Help

Address http://www.nursa.org/comp.cfm?detail=assay&dataset=25&cx_no=25.1

Are you a first time visitor? Be sure to register

Datasets

CAPER

Home | CAPER | Datasets | Experiment Data

Investigator : Bert O'Malley
Dataset : Proteomic Analysis of Steady-State Nuclear Hormone Coactivator Complexes
Summary : This dataset is an affinity purification-based profile of for nuclear receptor coactivators
Project : Proteomic Profiling of Coactivators
Technique : Affinity Purification
Molecule : CAPER
Antibody : BL462
Cell Source : Normal HeLa Cell nuclear extract, 15 mg total protein
Experiment Name : Caper, nuclear extract

Band	Name
1	[illegible]
2	Acinus [illegible]
3	[illegible]
4	SRRM1 [illegible]
5	[illegible]
6	U2AF2 [illegible]
7	PCBP3 [illegible]
8	SFR7 [illegible]
9	SAP18 [illegible]
10	SnRNP core D1 (SNRPD1)
	ASF 3 (SFRS1)

Notes

Molecules marked * are possible nonspe
More information can be found here.

Citation

Sung Yun Jung, Anna Malovannaya, Bert
Proteomic Analysis of Steady State Nucl
Datasets of the Nuclear Receptor Signaling

© 2005 - 2005 NURSA, The Nuclear Receptor Signaling

NURSA | Nuclear Receptor Signaling Atlas | Receptors, Coactivators, Corepressors and Ligands

Molecules

SnRNP core D1 (SNRPD1)

At a Glance | Gene Information | Protein Information | Other Information

Description : Small nuclear ribonucleoprotein Sm D1 (snRNP core protein D1) (Sm-D1) (Sm-D autoantigen).

Structure

Header	Source	Type
RNA BINDING PROTEIN	HOMO SAPIENS	X-RAY DIFFRACTION

PDB

The Worldwide Repository for Biological Macromolecular Structures

Structure Summary Page

Title : CRYSTAL STRUCTURE OF THE D1D2 SUB-COMPLEX FROM THE HUMAN SNRNP CORE DOMAIN

Figure 2. Hyperlinking the research community. Screenshots taken from www.nursa.org/ (status July 2005) indicating efficient, intuitive and intelligent browsing. Curated primary data (large screen) is cross-linked to far-reaching generic information: basic data housed within the NURSA domain (top right) and hyperlinked content of an external database (PDB in this example, bottom right). Primary NURSA data shown are peptides co-immunoprecipitated by immobilized estrogen receptor (ESR1, NR3A1)–coactivator CAPER (RNPC2) in HeLa cells (12).

- NURSA-curated Molecule Pages;
- NURSA laboratory web pages; and
- submissions to *NRS*, NURSA's open access journal.

Curated and generic data is accessible to individual users at www.nursa.org, as well as being exported to linked databases such as PubMed Central. We are committed to adopting emerging community standards for data exchange, such as the Extensible Markup Language (XML)-based frameworks under development for microarray and molecular interaction data (8,9). Finally, a central editorial and curatorial team exercises overall control over the submission, annotation and distribution of data to and from the website.

NURSA experimental data

NURSA laboratories are developing accelerated throughput initiatives in applications as diverse as microarray-based global transcriptomics, real-time quantitative PCR (Q-PCR) based gene expression analyses (10,11), proteomic and kinomic profiling of cellular proteins (12) and ligand screening (13). Other NURSA initiatives such as siRNA applications in high-content screening and ChIP-on-Chip assays have recently been initiated. A central goal of the Bioinformatics Resource is to integrate these primary datasets to construct reciprocal models of the interplay between events at the molecular level (ligand, receptor and coregulator modulation of target gene expression) and functionality at the organism level (organ physiology, metabolic regulation and homeostasis). Implicit in this objective is the broader goal of the consortium to develop improved translational therapies for intervention in the numerous diseases in which these molecules are implicated.

External data

External data repositories are mined with the aim of extracting and presenting information from these databases relevant to nuclear receptor, their ligands and coregulators. This information is used to support primary data derived from the NURSA laboratories by presenting users with relevant contextual information, as well as to provide accurate cross-references to appropriate content in external databases (as shown in Figure 2). These resources include nucleic acid-based repositories such as the NCBI database collection (with RefSeq, GenBank and UniGene), the NCBI's literature resource PubMed/MEDLINE, protein-based repositories such as the UniProt/EBI-databases SwissProt and TrEMBL, NucleaRDB, PIR, InterPro, GOA, Pfam and the protein structure database PDB, the compound databases PubChem, ChEBI and KEGG, and portals for species-specific gene collections such as MGI (for mouse data), RGD (rat), SGD (yeast), WormBase (*Caenorhabditis elegans*) and FlyBase (*Drosophila*).

NURSA-curated molecules

Molecular data for nuclear receptor ligands, coregulators and nuclear receptor orthologs from public repositories and literature databases is NURSA-internally curated and cross-linked to a content management system built upon the unique identifiers for molecules, assays and tissue systems. This ensures that the user is provided with appropriate contextual information. For example, a page displaying a dataset for estrogen receptor contains links to all other NURSA datasets and journal articles relevant to that molecule.

We have manually gathered to date ~2000 sequences of each nuclear receptor and coregulator orthologs from a large variety of species, and will automate the identification and parsing of sequences in the next phase of development to ensure that only updated information is made available to the community. Data for the nuclear receptors are organized by phylogeny to make the information adhere to the suggested unified nomenclature system for the NR Superfamily (14). Coregulators, however, which lack common sequence or structural features, are currently being classified based on functional properties, along with a variety of identifiable protein motifs found in other known molecules.

For the nuclear receptor ligands, publicly available data (e.g. data from ChEBI, KEGG and PubChem) have been augmented in a collaboration with GlaxoSmithKline, which has provided practical, useful information such as recommended concentration ranges for ligand studies.

Nuclear receptor signaling

Our PubMed-indexed journal *Nuclear Receptor Signaling* (*NRS*) offers members of the community an opportunity to contribute to the development of the NURSA website in the familiar, traditional form of a citable journal article. *NRS*'s article architecture is based entirely on the current XML Journal Publishing Document Type Definition created by PubMed Central, the U.S. National Library of Medicine's permanent digital archive of biomedical and life sciences journal literature. The effort invested in designing a PubMed Central-compliant electronic journal will, we anticipate, be repaid once *NRS* articles are listed on PubMed. About 60 million searches a month are run through the PubMed web interface, and having *NRS* content indexed in this database provides a portal to the NURSA website that would otherwise not exist. Consistent with NURSA's commitment to emerging community standards in electronic publishing, we have adopted the Digital Object Identifier (DOI) system (http://www.crossref.org/), and the Budapest Open Access Initiative (OAI; http://www.soros.org/openaccess/). The DOI system is designed to endow digital content with persistence and tangibility associated with a physical reproduction of this content—such as a printed journal article. OAI articulates the commitment of its signatories to the principle of free, less restrictive access to scientific data for the entire community.

DATA MINING AND WEBSITE DESIGN

For a website that seeks to establish a broad user base, the diversity of the nuclear receptor community—basic science, clinical therapy, pharmaceutical research and development, and technology transfer agencies—presents its own problems for the design and functionality of the site. Accordingly, we set out to develop a user interface based upon fundamental principles of web navigation, incorporating navigational elements that would provide for efficient, intuitive and intelligent browsing and searching of site content. Figure 2 illustrates how users are given the option, when viewing a primary NURSA dataset, to explore relevant information in external databases using numerous hyperlinks. In this example, proteins co-immunoprecipitated by immobilized estrogen receptor (ESR1, NR3A1)–coactivator CAPER (RNPC2) are shown by presenting the raw data (gel picture) as well as by listing the identified proteins. These proteins are linked into a large pool of generic data for the selected molecule, and, with only one more mouse click, the user can access the information in the external repository (as shown with the structure of the SNRPD core domain 1 at PDB in this case).

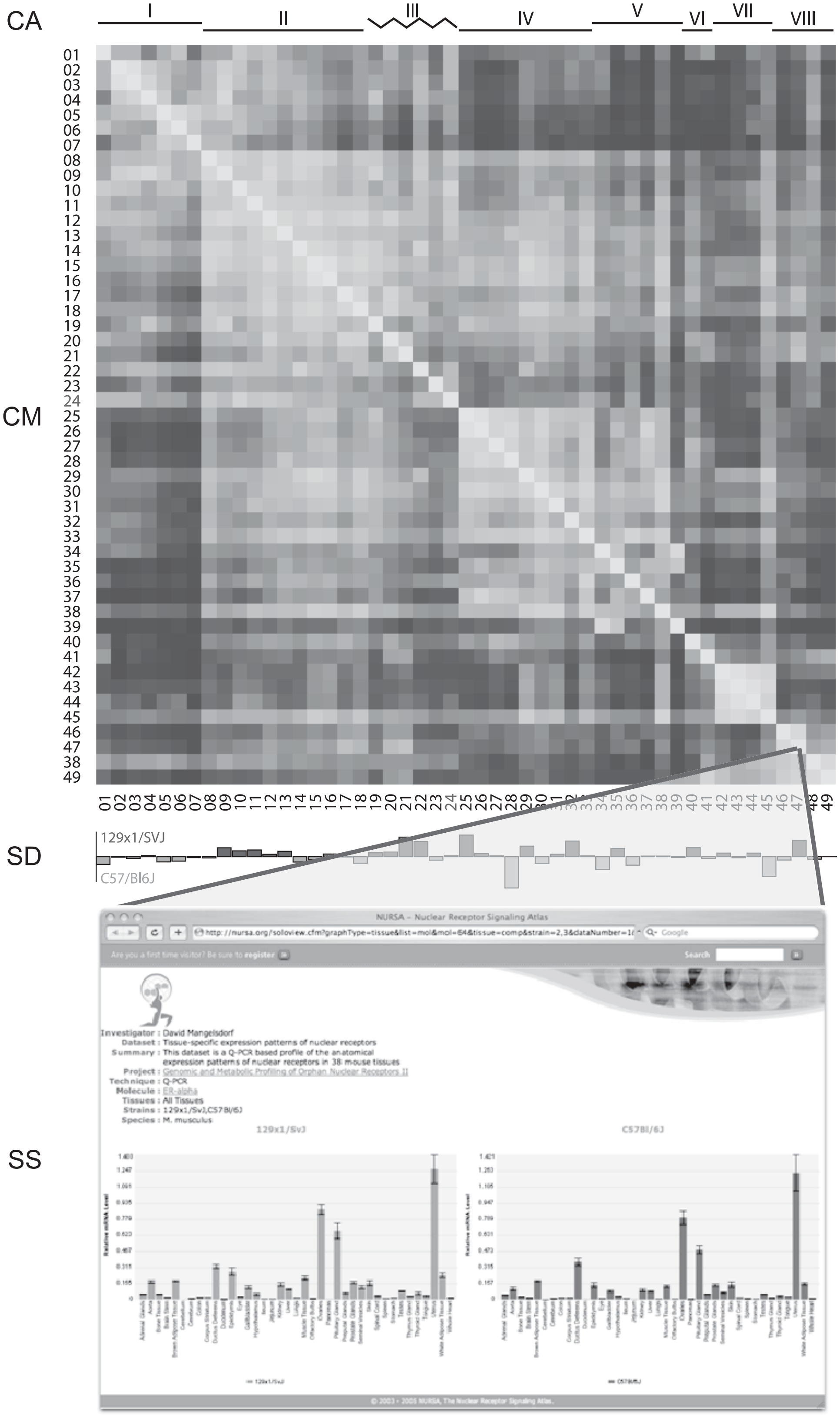
In addition to presenting primary data, the resource is also developing approaches to finding patterns, trends and relationships in the NURSA datasets that would not be apparent to the user when viewing individual experiments. The net effect of such efforts will be to work discrete datasets into a collective framework for making decisions, confirming directions and validating approaches—effectively placing the user on another level with respect to understanding the data.

To illustrate this with an example of what this means for the user in practical terms, Figure 3 shows a data analysis model we developed for the quantitative real-time PCR-based survey of the anatomical expression profiles of 49 nuclear receptors in 39 tissues in two mouse strains (15). Our objectives with regard to this dataset were to allow for the identification of similar expression profiles and to reveal unique nuclear receptor/tissue distribution patterns using objective analytical clustering combined with data correlation matrices. Each correlation matrix represents clustered nuclear receptor expression data from the two mouse strains analyzed in superimposed arrays. The normalized expression data from the 129x1/SvJ strain was color-coded in red and arranged in perpendicular arrays; C57Bl/6J results were color-coded green and also arranged as a matrix, and both square arrays subsequently superimposed to form a single correlation matrix. Because data were computed for nuclear receptors as well as for tissues, two distinct correlation matrices are generated, one for nuclear receptor and one for tissues (Figure 3 shows the matrix for nuclear receptors).

In such a correlation matrix, the yellow color is the result of equal distribution of red and green data, indicating comparable relative levels of expression in both strains. Color shifts towards green or red indicate predominance of either of the two strains. Besides the deviations in the expression pattern between strains, the matrix also gives a perspective on the relative expression levels, since we plotted arrays of normalized expression values to render the color intensity of each element within the matrix proportional to the strength of the correlation. In practical terms, the matrices give the user, in a single illustration, an overall 'high-level' perspective of the entire Q-PCR dataset. Because matrix annotations have been linked to a side-by-side display of the expression profiles of the selected nuclear receptor, such 'at-a-glance' representations of comprehensive information allow for a rapid data access. The user is again provided with multifaceted information and access options via an intuitive interface on the NURSA website.

CONCLUSION

We have described here a suite of features, both existing and under development, which are designed to add value to and enrich the experience of members of the nuclear receptor community. The concept of a specialist nuclear receptor database such as NURSA is predicated on the argument that such initiatives, with a more focused compass than larger generic resources, as well as a closer relationship with their respective fields of expertise, are positioned to provide a level of annotation that is not practicable for larger databases. Whether this argument is valid will ultimately be tested by the response of the community to the NURSA website and its content. Ultimately, our goal is to provide a platform for organizing and integrating the diverse data sources of this field into a framework to allow for better approaches to understanding the biology of nuclear receptors.



Investigator : David Mangelsdorf
Dataset : Tissue-specific expression patterns of nuclear receptors
Summary : This dataset is a Q-PCR based profile of the anatomical expression patterns of nuclear receptors in 38 mouse tissues
Project : Genomic and Metabolic Profiling of Orphan Nuclear Receptors II
Technique : Q-PCR
Molecule : ER-alpha
Tissues : All Tissues
Strains : 129x1/SvJ,C57Bl/6J
Species : M. musculus

Figure 3. Data mining efforts for integrated solutions. Combined correlation matrix (CM) for the Q-PCR anatomical expression profiling dataset for mouse nuclear receptors along with the results of cluster analysis (CA) and data integration (SD) allow for 'at-a-glance' interpretations of a comprehensive dataset as well as rapid data access. The later is illustrated by a screenshot (SS), which shows a side-by-side representation of the expression profiles of ER-alpha in 39 tissues of 129x1/SvJ and C57Bl/6 mice (see text for details).

ACKNOWLEDGEMENTS

This work was supported by a cooperative agreement funded through grant U19 DK62434 from the NIH. Funding to pay the Open Access publication charges for this article was provided by NIDDK.

Conflict of interest statement. None declared.

REFERENCES

1. Tsai,M.J. and O'Malley,B.W. (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem.*, **63**, 451–486.
2. Evans,R.M. (1988) The steroid and thyroid hormone receptor superfamily. *Science*, **240**, 889–895.
3. Rosenfeld,M.G. and Glass,C.K. (2001) Coregulator codes of transcriptional regulation by nuclear receptors. *J. Biol. Chem.*, **276**, 36865–36868.
4. McKenna,N.J., Lanz,R.B. and O'Malley,B.W. (1999) Nuclear receptor coregulators: cellular and molecular biology. *Endocr. Rev.*, **20**, 321–344.
5. Margolis,R.N., Evans,R.M. and O'Malley,B.W. (2005) The nuclear receptor signaling atlas: development of a functional atlas of nuclear receptors. *Mol. Endocrinol.*, **19**, 2433–2436.
6. Date,C.J. (2004) *An Introduction To Database Systems. 8th edn*. Pearson/Addison Wesley, Boston.
7. Codd,E.F. (1998) A Relational Model of Data for Large Shared Data Banks. 1970. *MD Comput.*, **15**, 162–166.
8. Brazma,A., Hingamp,P., Quackenbush,J., Sherlock,G., Spellman,P., Stoeckert,C., Aach,J., Ansorge,W., Ball,C.A., Causton,H.C. *et al*. (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genet.*, **29**, 365–371.
9. Hermjakob,H., Montecchi-Palazzi,L., Bader,G., Wojcik,J., Salwinski,L., Ceol,A., Moore,S., Orchard,S., Sarkans,U., von Mering,C. *et al*. (2004) The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.
10. Barish,G.D., Downes,M., Alaynick,W.A., Yu,R.T., Ocampo,C.B., Bookout,A.L., Mangelsdorf,D.J. and Evans,R.M. (2005) A Nuclear Receptor Atlas: Macrophage Activation. *Mol. Endocrinol.*, **19**, 2466–2477.
11. Fu,M., Sun,T., Bookout,A.L., Downes,M., Yu,R.T., Evans,R.M. and Mangelsdorf,D.J. (2005) A Nuclear Receptor Atlas: 3T3-L1 Adipogenesis. *Mol. Endocrinol.*, **19**, 2437–2450.
12. Jung,S.Y., Malovannaya,A., Wei,J., O'Malley,B.W. and Qin,J. (2005) Proteomic analysis of steady-state nuclear hormone receptor coactivator complexes. *Mol. Endocrinol.*, **19**, 2451–2465.
13. Downes,M., Verdecia,M.A., Roecker,A.J., Hughes,R., Hogenesch,J.B., Kast-Woelbern,H.R., Bowman,M.E., Ferrer,J.L., Anisfeld,A.M., Edwards,P.A. *et al*. (2003) A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell*, **11**, 1079–1092.
14. Nuclear Receptors Nomenclature Committee. (1999) A unified nomenclature system for the nuclear receptor superfamily. *Cell*, **97**, 161–163.
15. Bookout,A.L. and Mangelsdorf,D.J. (2003) Q-PCR Tissue-specific expression signatures of nuclear receptors. Datasets of the Nuclear Receptor Signaling Atlas. *Nuclear Receptor Signaling*, www.nursa.org, DOI: 10.1621/datasets.01006.